PROGRAM SYLLABUS

MASTER Data Science

Quarter 1
- Systems for Big Data Analytics
- Big Data Framework
- Convex Analysis and Optimization Theory
- Practical Introduction to Machine Learning
- An Introduction to Machine Learning Theory
- Statistical Learning Theory
- High Dimensional Statistics
- Non Parametric Estimation and Testing
- Probabilistic Graphical Models - Applications to Information Access
- Markov Chain Monte Carlo- Theory and Practical Applications
- Hidden Markov models and Sequential Monte Carlo methods
- Natural Language Processing and Sentiment Analysis
- Deep Learning I
- Data-Based Generative Models

Quarter 2
- Data Stream Processing
- Optimization for Data Science
- Non Differentiable Optimization and Proximal Methods
- Generalisation properties of Algorithms in ML
- High Dimensional Matrix Estimation
- Causal Inference
- Online Learning and Aggregation
- Advanced AI Methods for Graphs and NLP (ALTEGRAD)
- Introduction to Computer Vision

Introduction to Reinforcement Learning
Partially Observed Markov Chains in Signal and Image
Law and Ethics of Artificial Intelligence
Machine Learning : Business cases
Operations Research and Big Data
Cloud Data Infrastructure
Auction Theory and In Practice
Deep Learning Advanced
Introduction to Time Series
A Mathematical Introduction to Compressed Sensing
Deep Learning II
Audio and Music Information Retrieval
Mixed Effects Model : Methods, Algorithms, and Applications in Life Sciences
NLP
Applied Deep Learning with Python
Big Data and Insurance Project
Structured Data
Tail Events Analysis : Robustness, Outliers and Models for Extremes
Topics in Stochastic Filtering, Information and Control
Optimal Transport : Theory, Computations, Statistics, and ML applications

QUARTER 1

Systems for Big Data Analytics

Instructor: Angelos Anadiotis

Credits: 3 ECTS

Grading: It will be an in class examination which will consist in theoretical questions ⅓ and computer experiments ⅔.

Language: English

Syllabus: Modern data science applications require computer or domain scientists to work with various types of data, of big volume and which may be created at different rates: essentially, what we call Big Data. The efficiency of the analysis, in terms of both accuracy and speed, is critical for the applications, which may span across scientific experiments, finance and high frequency trading, manufacturing and many more. To improve speed, we often rely on large-scale infrastructures provided by data centers, where we deploy our algorithms. However, the use of such infrastructures require specialized software and careful algorithmic design to fully exploit the power of the hardware infrastructure and optimize for the cost of its allocation. This course considers the software and the hardware jointly and studies different angles of data analysis algorithms with respect to their performance over different types of modern and widely used hardware. It relies on fundamental aspects of data processing and it visits them under the light of different properties of the hardware and the application workload.

Objectives:
The objective of this course is to understand how to define and use a mechanistic model. Through examples (PKPD, viral dynamics,...) we will study the methods and algorithms used for mixed-effects models: methods of parameter estimation, model construction, validation and selection.

R and the Monolix software (http://lixoft.com/products/monolix/) will be (widely) used in the course

Format : 6 sessions of 3.5 hours + Exam

References:

- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.
- Verbeke, G. and Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. Springer.
- Commenges, D., & Jacqmin-Gadda, H. (2015). *Dynamical biostatistical models* (Vol. 86). CRC Press.
- Lavielle, M. (2014). *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press.

Big Data Framework

Instructor: Duc Pham Hi

Credits: 5 ECTS

Grading: The final mark of the module is a weighted average of 3 marks:

- Big Data with Hadoop (weight 5)
- Data Science with Spark (weight 3)
- Real Time Big Data Search and Analytics with Elastic Search (weight 2)

The first course is evaluated by an exam (1 hour) and a continuous evaluation (labs and mini-projects). The 2 last courses have a final exam (or a project, depending on students' orientations) only.

Numerus clausus: 30

Language: English

Syllabus: The objectives of this course are the following:

- Discover the different components of a Big Data cluster and how they interact.
- Understand Big Data paradigms.
- Understand the benefits of open source solutions.
- Develop a Big Data project from scratch
- Master Spark, its data models and its different methods of operation
- Learn how to use Spark to analyze data, develop Machine Learning pipelines and finally do streaming with Spark
- Understand and implement distributed algorithms.
- Understand the advantages of SQL/NOSQL databases.

Description of the course:
The module Big Data Frameworks is composed of three courses :

- Big Data with Hadoop (5 x 4 hours)
- Data Science with Spark (3 x 4 hours)
- Real Time Big Data Search and Analytics with Elastic Search (2 x 4 hours)

Big Data with Hadoop:

Apache Hadoop has been evolving as the Big Data platform on top of which multiple building blocks are being developed. This course presents the Hadoop ecosystem, Hadoop Distributed File System (HDFS) as well as many of the tools developed on it:

- MapReduce and YARN
- Hive and HBase
- Kafka, Flume, NiFi, Flink, Oozie, etc.

Students will also discover various subjects such as security, resource allocation and data governance in Hadoop.

Data Science with Spark:

Apache Spark is rapidly becoming the computation engine of choice for big data. This course presents:

- Spark's architecture and Spark Core: RDDs (Resilient Distributed Datasets), Transformations, and Actions
- Spark and Structured Data: explore Spark SQL and Spark Data Frames
- Spark Machine Learning libraries (MLLIB and ML)
- Spark Streaming

Real time Elastic search and Analytics

Powerful tool to analyze in real time tons of data available to provide insights to business.

Use cases:

- Search (Wikipedia, Github), Logging Analytics (Blizzard), Security Analytics (Slack), Metrics Analytics (NASA), Business Analytics (Tinder)
- High availability and Speed
- Elastic Stack:
- Collect (Beat)
- Ingest (Logstash)
- Store, search and analytics engine (Elasticsearch)
- Visualize (Kibana)
- The secret sauce of ES, Index, Shards (Primary, Replicas), Nodes, Cluster, Cross-Cluster
- Typical deployment in key use cases (search, logging, security)

Prerequisites:

Java, Python, Machine Learning and basic knowledge in Linux system administration and SQL.

Convex Analysis and Optimization Theory

Instructor: Pascal Bianchi, Olivier Fercoq

Credits: 5 ETCS

Grading: Final exam

Language: English

Syllabus: The course is not intended to provide a repertoire of algorithms as abundant as possible. It is about taking a step back in order to understand the mathematical foundations for the construction of a large class of iterative methods. The first part of the course is about convex analysis. We shall review the properties of convex functions, Fenchel-Legendre transform, and introduce the student to the duality theory in convex optimization. The second part of the course is about numerical algorithms. We shall see the conditions under which we can demonstrate the convergence of fixed point algorithms. This general approach makes it possible to obtain, as a corollary, the convergence of several emblematic algorithms such as the proximal gradient algorithm, or primal-dual algorithms. These algorithms are frequently used to solve optimization problems involving complex and structured regularizations, or optimization problems under constraints. They are frequently encountered in statistical learning, signal processing, and image processing. The objectives of the course are:

- Master the mathematical tools for the construction of optimization algorithms.
- Know how to demonstrate the convergence of iterates.
- Know how to numerically solve optimization problems involving non-differentiable terms and/or structured regularization.

Practical Introduction to Machine Learning

Instructor: Rémi Flamary

Credits: 3 ECTS

Grading : Practical session reports and oral

Numerus clausus: 40

Language : English

Syllabus : The objective of this course is to provide a practical introduction to the field of machine learning. We will discuss the different machine learning problems from unsupervised (dimensionality reduction, clustering and density estimation) to supervised (classification, regression, ranking). In this course we will introduce for each method the problem, provide its modeling as an optimization problem and discuss the algorithms that are used to solve the problem. The practical aspect of each method will also be discussed along with python code and existing implementations.

The course will be completed by practical sessions that will allow the students to implement the methods seen in the course on practical problems such as image classification and time series prediction (biomedical and climate data). The objective of the practical session will be not only to learn to use the methods but also to interpret their models and results with respect to the data and the theoretical models.

Course overview:

- Introduction
  - Machine learning problems
  - Knowing your data
  - Preprocessing
- Unsupervised learning
  - Dimensionality reduction and
  - Dictionary learning and collaborative filtering
  - Clustering and generative modeling
  - Generative modeling
- Supervised learning
  - Linear models and kernel methods for regression and classification
  - Nearest neighbors and bayesian decision
  - Trees and ensemble methods
- ML in practice
  - Find your problem
  - Model selection

An Introduction to Machine Learning Theory

Instructor: Stephan Clémençon, Myrto Limnios

Credits: 3

Language: French

Syllabus: Beaucoup d'applications modernes (génomique, finance, e-marketing, etc.) requièrent de manipuler et traiter des données de très grande dimension. La discipline qui développe et étudie des méthodes concrètes pour modéliser ce type de données s'appelle l'apprentissage statistique («statistical machine-learning»). Il s'agit, in fine, de produire des outils de prédiction et d'aide à la décision dédiés à une application spécifique. L'apparition d'algorithmes très performants pour la classification de données en grande dimension, tels que le boosting ou les Support Vector Machines dans le milieu des années 90, a progressivement transformé le champ occupé jusqu'alors par la statistique traditionnelle qui s'appuyait en grande partie sur le prétraitement réalisé par l'opérateur humain. En s'appuyant sur la théorie popularisée par Vapnik (The Nature of Statistical Learning, 1995), un nouveau courant de recherche est né: il se situe à l'interface entre les communautés mathématique et informatique et mobilise un nombre croissant de jeunes chercheurs tournés vers les applications liées à l'analyse de données massives. Dans ce module, on présentera le domaine, ses fondements théoriques, les problèmes qu'il permet d'aborder (apprentissage supervisé/non supervisé, batch/online, par renforcement, multi-tâche, asynchrone, etc.) et les approches algorithmiques les plus populaires.

«Nothing is more practical than a good theory» - V. Vapnik

Format: 6 sessions of lessons/practical lessons of 3h (1h30+1h30)

L'objectif du cours est de découvrir les enjeux et paradigmes du "machine learning", une discipline en plein essor à l'interface des mathématiques (probabilités/ statistiques, optimisation) et de l'informatique et qui joue aujourd'hui un rôle majeur en matière d'innovation technologique. Il s'agira ici d'en explorer quelques concepts et techniques essentiels, principalement autour du problème fondamental de la "classification supervisée" (i.e. "reconnaissance de formes"). Il se déroulera sur six séances de 3h incluant:

• une partie 'cours magistral' lors de laquelle seront formulés les problèmes et décrites certaines solutions de l'état de l'art ;
• une partie 'travaux dirigés' pour les séances d'exercices.

Séance 1 - 20/09:

• Introduction générale du cours : repères historiques, enjeux, applications, nomenclature des problèmes
• Le problème de la classification binaire (reconnaissance de formes) : Formalisme – Optimalité
Lectures conseillées: Chapitre 2 de (1), Chapitres 1 et 2 de (9), article (4)

Séance 2 - 27/09

• Théorie probabiliste de la classification - Minimisation empirique du risque
• Théorie de Vapnik-Chervonenkis – Complexité combinatoire - Moyennes de Rademacher
• Exercices Lectures conseillées: articles (3) et (4)

Séance 3 - 04/10

• Premières stratégies d'apprentissage supervisé, modélisation et moyennes locales: régression logistique - perceptron - arbres de classification – K-plus proches voisins - réseaux de neurones Lectures conseillées: Chapitres 4 et 9 de (1)

Séance 4 - 11/10

• Evaluation de l'erreur et sélection de modèles : plan expérimental – bootstrap – validation croisée – minimisation structurelle du risque • Ensemble Learning: Bagging, Boosting et Forêts Aléatoires Lectures conseillées: Chapitre 7 de (1)

Séance 5 - 18/10

• Les machines à vecteurs support (SVM) : linéaires/non linéaires
• «Kernel trick»: ACP, régression Lectures conseillées: (8) et (9) Séance 6 - 25/10
• Au delà des problèmes d'apprentissage 'locaux' (classification, regression, estimation de densité): clustering, ranking, détection d'anomalies

References: Les «slides» du cours seront disponibles en version électronique. On se réfèrera en particulier aux documents suivants.

•Friedman, Hastie & Tibshirani (2009). The Elements of Statistical Learning. Third edition, Springer. Disponible en ligne.
•Bousquet, Boucheron & Lugosi (2004). Introduction to statistical learning theory. In O. Bousquet, U.V. Luxburg, G. Rätsch (editors), Advanced Lectures in Machine Learning, Springer, pp. 169-207, 2004. Disponible en ligne.
•Bousquet, Boucheron & Lugosi (2004). Concentration Inequalities. In Advanced Lectures in Machine Learning, Springer, pp. 208-240. Disponible en ligne.
•Kulkarni, G. Lugosi & S. Venkatesh (1998). Learning Pattern Classification. A Survey. 1948-1998 Special Commemorative Issue of IEEE Transactions on Infor mation Theory, vol.44, 2178-2206. Reprinted in S. Verdú, S.W. McLaughlin (editors.), Information Theory: 50 Years of Discovery, IEEE Press, New York, 1999. Disponible en ligne.
•Cesa-Bianchi & Lugosi (2006) Prediction, Learning, and Games. Cambridge. University Press.
•Devroye, Györfi & Lugosi (1996) A Probabilistic Theory of Pattern Recognition. Springer
•Györfi, Kohler, Krzyzak & Walk (2002) A Distribution-Free Theory of Nonparametric Regression. Springer
•Burgess. A Tutorial on SVM for Pattern Recognition. Kluwer. Disponible en ligne.
•Vapnik. The Statistical Nature of Learning Theory. Springer.

Statistical Learning Theory

Instructor: Jaouad Mourtada (ENSAE)

Credits: 3 ects

Grading: Final exam. An exercise sheet (with exercises in the spirit of the exam) will be provided, and will be partly corrected in class.

Numerus Clausus: 50

Language: English

Syllabus: Supervised statistical learning refers to the following prediction problem: find a good rule to predict some output/label of interest (for instance, the object represented by an image), based on some associated input/features (for instance, the image itself), using a dataset of feature-label pairs. This is a core machine learning problem, with applications to computer vision, speech recognition and natural language processing, among many domains.
The aim of this course is to provide an introduction to the theory and principles underlying this general problem. The emphasis will be placed on mathematical and conceptual aspects, rather than on actual implementation: the key concepts and ideas will be illustrated through formal (often elementary) results, which will be proven in class.
Along the way, we will introduce some technical tools relevant in statistics and machine learning theory, including basic concentration inequalities and control of empirical processes, stochastic gradient methods, and approximation properties of some classes of functions.

Topics covered:

- Probabilistic setting: prediction, loss and risk;
- Universal consistency and No Free Lunch theorem;
- Empirical risk minimization and its analysis via uniform convergence. Approximation and estimation errors, overfitting and underfitting;
- Finite classes and Hoeffding's inequality; Rademacher complexity and Vapnik-Chervonenkis classes (upper and lower bounds);
- Model selection: complexity regularization and cross-validation;
- Nonparametric regression by histograms, bias-variance decomposition and convergence rates over Hölder classes; curse of dimension;
- Convex surrogate losses and their properties; Linear prediction, stochastic gradient methods; Links between regularization and optimization;
- Learning in high dimension and neural networks: approximation of 2-layer networks, interpolation and overfitting for linear regression.

Prerequisites: Knowledge of probability theory at an undergraduate level. Some notions in statistics (at the level of a first course in mathematical statistics) would be helpful, though this is not a strict requirement. Some general background in machine learning would also be a plus, though none is required to follow this course.

References:

Lecture notes will be provided before each class.
-- S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: A survey of some recent advances. ESAIM: probability and statistics, 9:323–375, 2005. Link
-- L. Devroye, L. Györfi, and G. Lugosi. A Probabilistic Theory of Pattern Recognition, volume 31 of Applications of Mathematics. Springer-Verlag, 1996. Link
-- A. B. Tsybakov. Introduction to nonparametric estimation. Springer, 2009.
-- L. Györfi, M. Kohler, A. Krzyzak, and Harro Walk. A distribution-free theory of nonparametric regression. Springer Science & Business Media, 2002. Link
-- S. Shalev-Shwartz and S. Ben-David. Understanding machine learning: From theory to algorithms. Cambridge University Press, 2014. Link

High Dimensional Statistics

Instructor: Alexandre Tsybakov

Credits: 3 ects

Grading: Final exam

Numerus clausus: 15

Language: English

Syllabus: This course develops tools to analyze statistical problems in high-dimensional settings where the number of variables may be greater than the sample size. It is in contrast with the classical statistical theory that focuses on the behavior of estimators in the asymptotics as the sample increases while the number
of variables stays fixed. We will show that, in high-dimensional problems, powerful statistical methods can be constructed under such properties as sparsity or low-rankness. The emphasis will be on the non-asymptotic theory underlying these developments.

Topics covered:
- Sparsity and thresholding in the Gaussian sequence model.
- High-dimensional linear regression: Lasso, BIC, Dantzig selector, Square
- Root Lasso. Oracle inequalities and variable selection properties.
- Estimation of high-dimensional low rank matrices. Matrix completion.
- Inhomogeneous random graph model. Community detection and estimation in the stochastic block model.

Prerequisites: Solid knowledge of probability theory, mathematical statis-
tics, linear algebra. Notions of convex optimization.

Resources:
Alexandre Tsybakov. High-dimensional Statistics. Lecture Notes.

Non Parametric Estimation and Testing

Instructor: Cristina Butucea

Credits: 3 ects

Grading: Final exam

Language: English

Syllabus:

Probabilistic Graphical Models - Applications to Information Access

Instructor: François Yvon (CNRS)

Credits: 3 ects

Grading: Final exam + homeworks

Numerus Clausus: 24

Language: English

Syllabus: Our information society produces an ever-increasing flow of unstructured data of various types (texts, audio, image, video, etc) that needs to be dealt with quickly and effectively. In the face of such polymorphic data, probabilistic models have emerged through their ability to digest the variability of information into effective and sound statistical models. In the last decades, these models have become indispensable tools for information management and decision-making. The course is divided into two main parts. The first part deals with the basic concepts and their computational manipulation: directed and undirected graphical models, and the associated algorithms. In the second part, we focus more specifically on (a) the estimation of latent variable models: (b) approximate inference. We will illustrate these methods with applications from the text mining literature (text classification and clustering, question answering, sentiment analysis, etc).

Main themes :

- Directed graphical model and probabilistic reasoning
- Undirected graphical model
- Exact inference in graphical models
- EM and latent variable models
- Approximate inference: variational techniques
- Approximate inference: sampling techniques

Prerequisites: Basic statistics and optimization

References:

Probabilistic Graphical Models: Principles and Techniques by Daphne Koller and Nir Friedman. MIT Press.
Pattern Recognition and Machine Learning by Chris Bishop.
Machine Learning: a Probabilistic Perspective by Kevin P. Murphy. MIT Press
Modeling and Reasoning with Bayesian networks by Adnan Darwiche. Information Theory, Inference, and Learning Algorithms by David J. C. Mackay. [Available online.]
Graphical models, exponential families, and variational inference by Martin J. Wainwright and Michael I. Jordan. [Available online]

Markov Chain Monte Carlo - Theory and Practical Applications

Instructors : Randal Douc & Sylvain Le Corff

Credits: 3 ECTS

Grading: Quiz, project based on a research article

Numerus Clausus: 30

Language: English

Syllabus: For the past few years, statistical learning and optimization of complex dynamical systems with latent data have been applied to time series analysis across a wide range of applied science and engineering domains such as signal processing, target tracking, enhancement and segmentation of speech and audio signals, inference of ecological networks, etc.
Solving Bayesian nonlinear filtering and smoothing problems, i.e. computing the posterior distributions of some hidden states given a record of observations, and computing the posterior distributions of the parameters is crucial to perform maximum likelihood estimation and prediction of future states of partially observed models. Estimators of these posterior distributions may be obtained for instance with Sequential Monte Carlo (SMC), also known as particle filtering and smoothing, and Markov Chain Monte Carlo (MCMC) methods.
This course sets the focus on MCMC algorithms and provides an overview of such approaches: introduction to standard procedures, convergence properties of a few algorithms and practical extensions (with simulations based on Python Notebooks) to more complex solutions.

Topics covered:

- Markovian models (specific focus on observation-driven models).
- Introduction to Markov chain Monte Carlo algorithms.
- Some convergence results of Markov chain Monte Carlo algorithms.
- Pseudo-Marginal MCMC and applications.
- Hamiltonian Monte Carlo algorithms and variants.
- Introduction to variational methods.

Hidden Markov Models and Sequential Monte Carlo Methods

Instructor: Chopin Nicolas

Credits: 3 ects

Grading : Project-based (group of three students)

Language: English

Syllabus: So-called hidden Markov chain (or state-space) models are time series models involving a "signal" (a Markov process (X_t) describing the state of a system) observed in an imperfect and noisy way in the form of data, e.g. $Y_t=f(X_t)+\epsilon_t$. These models are widely used in many disciplines:

- Finance: stochastic volatility (X_t is the unobserved volatility)…
- Engineering: target tracking (X_t is the position of a mobile whose trajectory we are trying to find; speech recognition (X_t is a phoneme).
- Biostatistics: Ecology (X_t=population size); Epidemiology (X_t=number of infected).

The aim of this course is to present modern methods of sequential analysis of such models, based on particle algorithms (Sequential Monte Carlo). The problems of filtering, smoothing, prediction, and parameter estimation will be discussed. At the end of the course, we will also briefly discuss the extension of such algorithms to non-sequential problems, notably in Bayesian Statistics.

At the end of the course, the student will be able to:

- state the main properties of HMM models
- implement a particle filter to filter and smooth a given HMM model
- estimate the parameters of such a model from different methods

Planning:

1. Introduction: definition of HMM (Hidden Markov models), main properties, notion of filtering, smoothing and prediction, forward-backward formulas.
2. Discrete HMMs, Baum-Petrie's algorithm
3. Gaussian linear HMM, Kalman algorithm
4. SMC algorithms for filtering an HMM model
5. Estimation in HMM models
6. Introduction to non-sequential applications of SMC algorithms

Prerequisites:

- 2A Simulation and Monte Carlo or similar course
- 3A courses of "Computational Statistics" and "Bayesian Statistics" are recommended but not mandatory.

References

Del Moral (2004). Feynman-Kac formulae, Springer; Chopin, N. and Papaspiliopoulos, O. (2020). An Introduction to Sequential Monte Carlo, Springer.

Natural Language Processing and Sentiment Analysis

Instructor : Chloé Clavel, Matthieu Labeau

Credits: 3 ects

Grading: Practical session reports (50%) and Reading note of a research paper (50%).

Numerus clausus: 40

Language: English

Syllabus: Many efforts have been recently dedicated to the development of methods able to analyze sentiment data available on the social Web or during interactions with conversational systems. The objective of this course is to tackle the different methods underlying natural language processing and sentiment analysis.

The techniques and concepts that will be studied include:

- Natural language pre-processing for sentiment analysis: tokenization, part-of-speech tagging, document representation and word embeddings techniques, natural language resources (ex. Sentiwordnet)
- Sentiment classification: from naïve to advanced machine/deep learning methods
- Sequential models for NLP: seq2seq and attention mechanisms

The course provides a mix between theoretical lectures and their practical labs where students learn to implement natural language processing and sentiment analysis framework using the following python libraries: pytorch, nltk and scikitlearn.

Format: 6 sessions (4 lectures and 2 Labs) of 3.5 hours

Schedule:

Lectures - Natural Language Processing methods for sentiment analysis – Part 1,2 &3 - Chloe Clavel
Lab - Sentiment analysis & embeddings – Matthieu Labeau
Lecture - Sequential Models for NLP - Matthieu Labeau
Lab - Sequential Model for NLP - Matthieu Labeau

Deep Learning I

Instructors: Geoffroy Peeters, Alasdair Newson (Télécom Paris, IP-Paris)

Credits: 3 ects

Grading : 30% labs/project + 70% written exam

Language: English

Syllabus: Deep Learning (machine learning based on deep artificial neural networks) has become extremely popular over the last years due to the very good results it allows for tasks such as regression, classification or genera- tion. The objective of this course is to provide a theoretical understanding and a practical usage of the three main types of networks (Multi-Layer- Perceptron, Recurrent-Neural-Network and Convolutional Neural Network). The content of this course ranges from the perceptron to the generation of adversarial images. Each theoretical lecture is followed by a practical lab on the corresponding content where student learn to implement these networks using the currently three popular frameworks: pytorch, tensorflow and keras.

Lectures content:

- Multi-Layer-Perceptron (MLP): Perceptron, Logistic Regression, Chain rule, Back-propagation, Deep Neural Activation functions, Vanishing gradient, Ini- tialization, Regularization (L1,L2,DropOut), Alternative Gradient Descent, Batch- normalization
- Recurrent Neural Network (RNN) : Simple RNN, Forward Propagation, Backward Propagation Through Time, Vanishing/ Exploding gradients, Gated Units (LSTM, GRU), Various architectures, Sequence-to-sequence, Attention model
- Convolutional Neural Network (CNN) : CNNs use sparse connectivity and weight sharing to reduce parameters and create more powerful networks , connections are organized in a convolution operation, CNNs now provide the state- of-the-art in a vast array of problems, we will see how CNNs work and we will implement them for classification problems

Labs content :

- text recognition, sentiment classification
- music generation
- image recognition
- image generation

Programming language

- Python (numpy, scikit-learn, matplotlib)
- DL frameworks: pytorch, tensorflow, keras
- Use Télécom computers, your own labtop or colab.research.google.com → needs a Google account → open one before the first Lab !
- a Graphics Processing Unit (GPU) will not be required, however if you have one, this will speed up the learning process

Data-based Generative Models

Instructor: Emmanuel Gobet

Credits: 3 ects

Language: English

Grading: written exam or project

Syllabus: A new paradigm of generative models has emerged in AI in the last decade. It aims at designing a generative model mimicking a distribution described by a data set. It has mainly two applications: data augmentation, i.e. to generate new data statistically coherent with those of the initial (training) data set; digital twin, i.e. to replace a costly physical simulation model with an easy-to-use one. This has huge applications in image generation, fashion pictures, chemical molecules... this is called sometimes deepfakes.

A generative model is usually made of a random input (the latent space) and of a deterministic transformation function (usually made of neural networks) mapping the latent space into the data space.
In this course, we will describe how to mathematically model and solve this problem, starting from basic concepts and going to more advanced tools. We will investigate (with mathematical tools)

- how to measure the distance between probability distributions, the relations between these distances
- existence of transformation between two given distributions
- the design of the optimization problem of the generative model
- possible parameterizations based on GAN, VAE, SDE
- role of the latent dimension in the search for a numerical solution, quality/complexity bounds on examples
- some statistical analysis for the sampling effect

QUARTER 2

Data Stream Processing

Instructor: Jérémie Sublime

Credits: 3 ects

Grading:
- The practical sessions will make ⅔ of the mark
- The research paper presentation will make ⅓ of the mark

Language: English

Syllabus: This course deals with the algorithms and softwares commonly used to process large data streams. It aims at understanding the main difficulties and specificities of this type of data, knowing what different types of streams exist, what are the theoretical models and practical algorithms to analyze them, and what are the right tools to process these streams.

After an introduction of what data streams are from a conceptual point of view, this class covers the question of data stream processing from two different angles:

1. A Machine Learning and Data Mining approach to cover the theoretical and algorithmic difficulties of learning from data streams: online learning vs incremental and batch learning, and sampling techniques.
2. A more practical approach with an introduction to the various systems and software that are used to handle these data.

In terms of organization, the course will consist of an alternance of lectures and practical sessions. Finally, during the last class the students will have to present a recent research article of their choice on the subject of data stream processing.

Objectives :
- Introduction to the concept of data stream processing
- Learning the basics on and how to use Data Stream Management Systems (DSMS)
- Understanding the main sampling techniques used for stream processing : sampling, sketching, etc.
- Understanding and using the main data stream processing algorithms

Prerequisites :
- Basics in SQL language
- Basics in Machine Learning (supervised and unsupervised)
- A knowledge of Java programming is recommended but not mandatory

Optimization for Data Science

Instructors: Alexandre Gramfort, Robert Gower

Credits: 5 ects

Grading:
Labs. 2-3 Labs with Jupyter graded (30% of the final grade).
Project. Implementation of solvers for a machine learning model. 30% of final grade.
Exam. 3h Exam (40% of the final grade).

Language: English

Syllabus : Modern machine learning heavily relies on optimization tools, typically to minimize the so called loss functions on training sets. The objective of this course is to cover the necessary theoretical results of convex optimization as well as the computational aspects. This course contains a fair amount of programming as all algorithms presented will be implemented and tested on real data. At the end of the course, students shall be able to decide what algorithm is the most adapted to the machine learning problem given the size of the data (number of samples, sparsity, dimension of each observation).

Planning:

1 Lecture [2.00 hours lecture + 1.50 hours exe, Robert] Foundations of convex optimization: gradient, sub-differential, strong convexity, conditioning. Examples with linear regression and logistic regression for which we compute gradients, Hessian, Lipchitz constants, etc. -Exercise list : Convexity and smoothness, Ridge regression and gradient descent

2 Lecture [2.00 hours lecture + 1.50 hours exe, Robert] First order algorithms: gradient descent, proximal operator, proximal gradient descent, convergence proofs in the smooth and smooth + strongly convex case. Give exercises. -Exercise list : Proximal Operator

3 Lab [3.50 hours Lab Robert + Alex + TAs] Lab (grad + prox grad) in jupyter notebooks. Lab to be graded. -Lab1.ipynb

4 Lecture [2.0 hours lecture + 1.50 hours exe, Robert] Stochastic algorithms, SGD, with and without moment convex, strongly convex, with proofs. -Exercise list : SGD and CD for ridge regression

5 Lecture [1.50 hours lecture + 2.0 hours Lab, Robert] Stochastic variance reduced methods (SAGA, SVRG) with numerical tricks: lazy updating, sparse tricks. - Lab SGD + variants (not graded) -Exercise list : Variance reduction

6 Lab [3.50 hours Lab Robert + Alex + TAs] Lab on all methods seen so far (GD, prox GD, Fast prox GD, SGD, SAG, SVRG, SDCA). Lab to be graded -Lab2.ipynb

7 Lecture [2 hour + 1.50 hours Lab, Robert] Online variance reduction and scale invariant methods (Natural gradient descent, ADagrad, Adam, randomized Newton)

8 Lecture [3.5 hours, Alex] Part I: Coordinate descent algorithms + Lab: coordinate descent implementation on logistic and ridge + Project introduction

9 Lecture [1 hour + 2.5 hours Lab, Alex] Solvers for quadratic functions (SVD, woodbury matrix inversion lemma), conjuguate gradient (dense vs sparse data). Line search methods.

10 Lecture [2 hours + 1.5 hours Lab, Alex] Second order methods: Newton, Quasi-Newton (BFGS, L-BFGS). Illustrate convergence problem of Newton, initialisation, construction of BFGS et L-BFGS.

11 Lecture [3.5 hours, Alex] non-convex and beyond convexity (adaptive Lasso / SCAD), Frank-Wolfe.

References:

Book 1. Boyd & Vandenberghe: Convex Optimization. Chapters 2, 3 and 4 for a revision on convexity and chapter 9 for a revision on unconstrained optimization. Freely available here.
Book 2. Shalev-Shwartz & Ben-David: Understanding Machine Learning, from Theory to Algorithms. Chapters 1 and 2 for a frequentist introduction to Machine Learning. Freely available here.
Book 3. Bubeck: Convex Optimization: Algorithms and Complexity. Chapter 6 for additional proofs for stochastic gradient methods including SGD and SVRG. Freely available here.
Paper 1. Amir Beck and Marc Teboulle (2009), SIAM J. Imaging Sciences, A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. Freely available here.
Paper 2. RMG et all (2019), Proceedings of Machine Learning Research, Volume 97, SGD: general analysis and improved rates freely available here.

Non Differentiable Optimization and Proximal Methods

Instructor: Andres Contreras

Credits: 3 ects

Grading: written exam

Language: English

Syllabus: A partir des outils de l'analyse convexe, l'objectif de ce cours est de présenter les algorithmes de résolution des problèmes d'optimisation non différentiables. Le cours fait appel à de nombreux exemples d'application et met en évidence la nécessité de prendre spécifiquement en considération le caractère non différentiable des problèmes.

La première séance est consacrée à l'exposé des principales propriétés des fonctions sous-différentiables, dans le cadre de l'analyse convexe. On détaillera les conditions d'optimalité générales dans le cas sous-différentiable, ainsi que les propriétés de différentiabilité des fonctions marginales en optimisation.
Les trois séances suivantes présentent plusieurs classes d'algorithmes en optimisation sous-différentiable, leurs applications dans le cadre de la relaxation Lagrangienne et leur utilisation en dualité : méthodes proximales et algorithmes du gradient proximal, méthodes de plans sécants et algorithme des faisceaux, méthode du recouvrement progressif en optimisation stochastique.

Prerequisites: convex analysis, subdifferentials, duality.

Generalization Properties of Algorithms in ML

Instructor: Aymeric DIEULEVEUT

Credits: 3 ects

Grading: Quizz + Présentation d'article / projet.

Numerus clausus: 40

Language: English.

Syllabus : La majorité des problèmes d'apprentissage sont formulés comme des problèmes d'optimisation, à partir de l'observation d'un échantillon de données (ensemble d'entraînement). L'optimisation d'un objectif défini à partir de cet échantillon permet de proposer un estimateur qui a une bonne performance sur l'ensemble d'apprentissage. Cependant, on s'intéresse généralement à la capacité de généralisation de cet estimateur, c'est-à-dire sa performance sur une nouvelle observation. Avec l'émergence des grandes quantités de données depuis les années 2000, le lien entre l'algorithme utilisé et la capacité de généralisation de l'estimateur associé est devenu un sujet majeur. Aujourd'hui, la question de la généralisation est encore une problématique de recherche majeure, tant pour ses aspects théoriques que pratiques. Dans ce cours, on s'intéresse à l'ensemble des résultats tant théoriques que heuristiques qui permettent d'aborder ce problème. Plus précisément, on étudiera dans un premier temps les différentes approches qui permettent d'obtenir des garanties théoriques quant à la généralisation des algorithmes, en particulier les approches liées à la complexité, à la stabilité et aux méthodes d'arrêt anticipé (Early stopping, approximation stochastique). Dans une seconde partie, on étudiera les approches heuristiques et les différences (expliquées ou constatées) dans le cadre du deep learning (non convexe et over-parametrized).

Prerequisites : connaissances élémentaires en optimisation convexe et statistiques. Avoir suivi le cours d'optimisation pour les data-sciences permettra de mieux cerner les différents algorithmes en jeu.

References :

- Rademacher and Gaussian Complexities: Risk Bounds and Structural Results, P. Bartlett, S. Mendelson
- The Tradeoffs of Large Scale Learning, L. Bottou, O. Bousquet - Stability and Generalization, O. Bousquet, A. Elisseef
- Train faster, generalize better: Stability of stochastic gradient descent, M. Hardt, B. Recht, Y. Singer
- Non-strongly-convex smooth stochastic approximation with convergence rate O(1/n), F. Bach, E. Moulines
- Understanding deep learning requires rethinking generalization, C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals
- On early stopping in gradient descent learning, Y Yao, L. Rosasco, and A. Caponnetto
- Generalization properties of multiple passes stochastic gradient method, S. Villa
- Competing with the empirical risk minimizer in a single pass, R. Frostig, R. Ge, S. M. Kakade, A. Sidford
- Deep Learning and Generalization, O. Bousquet

High Dimensional Matrix Estimation

Instructor: Karim Lounici

Credits: 3

Grading: final exam, article

Numerus Clausus: 30

Language: English

Syllabus : Nowadays many data learning problems require to analyze the structure of a high-dimensional matrix with remarkable properties; In recommender systems, this could be a column sparse matrix or a low-rank matrix but more sophisticated structures could be considered by combining several notions of sparsity; In graph analysis, popular spectrum techniques to detect cliques are based on the analysis of the Laplacian matrix with specific sparse/low-rank structure. In this course, we will review several mathematical tools useful to develop statistical analysis methods and study their performances. Such tools include concentration inequalities, convex optimization, perturbation theory and minimax theory.

Topics covered:

1. Principal Component Analysis
2. Spectral clustering
3. Matrix completion
4. Robust Statistics
5. Phase Retrieval
6. Optimal Transport

References:

R. Vershynin. High-Dimensional Probability. Cambridge University.

D. Gross, Recovering low-rank matrices from few coefficients in any basis, 2011, arXiv:0910.1879

O. Guedon and R. Vershynin. Community detection in sparse networks viagrothendieck's inequality.Probability Theory and Related Fields, 165(3-4):1025–1049,2016.

J. Ma, R. Dudeja, J. Xu, A. Maleki, X. Wang. Spectral Method for Phase Retrieval: an Expectation Propagation Perspective. arXiv: 1903.02505

W. M. Kouw, M. Loog. *An introduction todomain adaptation and transfer learning, 2018. arXiv:1812.11806*

Causal Inference

Instructor: Josse Julie

Credits: 3 ECTS

Numerus Clausus : 50

Grading: Exam (50%), Project (50%). It could be more practical with a data analysis or presentations based on recent research papers.

Syllabus : In machine learning, there has been great progress in obtaining powerful predictive models, but these models rely on correlations between variables and do not allow for an understanding of the underlying mechanisms or how to intervene on the system for achieve a certain goal. The concepts of causality are fundamental to have levers for action, to formulate recommendations and to answer the following questions: "what would happen if » we had acted differently?

The questions of causal inference arise in many areas (socio-economics, politics, psychology, medicine, etc.): depending on the context which drug to use to improve the patient's health? what marketing strategy for product placement should be used to influence consumer buying behavior, etc. The formalism of causal inference makes it possible to study these questions as a problem of classical statistical inference. The gold standard for estimating the effect of treatment is a randomized controlled trial (RCT) which is, for example, mandatory for the authorization of new drugs in pharmaceutical and medical research. However, RCTs are generally very expensive in terms of time and financial costs, and in some areas such as economics or political science, it is often not possible to implement an RCT, for example to assess the effectiveness of a given policy.

The aim of this course is to present the available methods to perform causal inference from observational data. We focus on both the theoretical framework and practical aspects (available software solution). In terms of application, the lecture will be illustrated with recent exemples mainly in the field of health : What is the effect of Hydrochloroquine on survival ? What would have happened if Italy's government had waited a week before imposing lockdown measures ? Ect.

Topics covered:
- The Neyman-Rubin potential outcome causal model for observational studies
- Matching, propensity scores
- Efficiency and double robustness, double machine learning
- Estimating treatment effect heterogeneity, causal forest
- Causal discovery : causal models, graphical models and markov conditions

References:
Hernan, Miguel A., and James M. Robins. Causal Inference. Chapman & Hall/CRC
Imbens, Guido W., and Donald B. Rubin. Causal Inference in Statistics, Social, and Biomedical Sciences
Jonas Peters, Dominik Janzing, Bernhard Schölkopf. Elements of Causal Inference: Foundations and Learning Algorithms
Articles : A Survey of Learning Causality with Data: Problems and Methods
Guo, Ruocheng and Cheng, Lu and Li, Jundong and Hahn, P. Richard and Liu, Huan.

Online Learning and Aggregation

Instructor: Alexandre Tsybakov

Credits: 3 ects

Grading: Final exam

Language: English

Syllabus: The aim of online learning is to provide efficient recursive algorithms of prediction when the data are arriving sequentially in a streaming way rather than as an array given once and for all. Whereas statistical learning is dealing with independent identically distributed data, the emphasis in online learning is on adversarial setting where the data are of arbitrary nature satisfying mild conditions. In this setting, one of the key ideas is to use, at each time instance, a suitable randomized choice from the given set of candidate predictors. Analogous techniques can be applied to solve the problem of aggregation, that is, to obtain procedures that predict almost as good as the best estimator in a given set. This course provides an introduction to online learning and aggregation focusing on theoretical aspects.

Topics covered:
-- Online classification in realizable case, halving.
-- Online gradient descent for convex and strongly convex loss. Online-to-batch conversion. Online linear regression.
-- Randomization by exponential weighting. Prediction with expert advice.
-- Adversarial multi-armed bandit problem.
-- Aggregation of estimators.
-- Gradient-free online learning. Continuum bandit problem.

References:

Shalev-Schwartz, S. (2011) Online learning and online convex optimisation. *Foundations and Trends in Machine Learning*, vol. 4, pages 107-194.
Tsybakov, A. (2020) Online learning and aggregation. Lecture Notes.

Advanced AI Methods for Graphs and NLP (ALTEGRAD)

Instructor: Michalis VAZIRGIANNIS

Credits: 5 ects

Grading: 20% from lab assignments and 80% from a data challenge

Language: English

Syllabus:

1. TEXT/NLP - Graph based Text Mining
- Graph-of-words GoWvis
- Keyword extraction (TFIDF, TextRank, ECIR'15, EMNLP'16)
- extractive summarization
- Sub-event detection in twitter streams
- graph based document classification: TW-IDF, TW-ICW, subgraphs abstractive summarization - ACL 2018 summarization

2. TEXT - NLP - Word & doc embeddings
- Word embeddings: word2vec-glove models, doc2vec, subword, Latent Semantic Indexing, context based embeddings
- doc similarity metrics: Word Mover's distance, shortest path kernels

3. Deep learning for NLP
- CNNs, RNNs LSTMs for NLP, text classification
- Meta-architectures
- Sequence to Sequence: Attention (HAN),
  - Domains: summarization
  - Translation, image captioning
- Unsupervised word sense detection/disambiguation
- French Linguistic resources: http://master2-bigdata.polytechnique.fr/FrenchLinguisticResources/

4. Graph kernels, community detection

Grakel python library: **https://github.com/ysig/GraKeL/**

5. Deep Learning for Graphs - node classification
- node embeddings (deepwalk & node2vec) for node classification and link prediction
- Supervised node embeddings (GCNN, ...)

6. Deep Learning for Graphs - Graph classification, GNNs
- graph CNNs
- message passing
- Graph - Auto-encoders

7. Sets embeddings - point clouds

8. Network Architecture Search - interpretability.

Course page: https://moodle.lix.polytechnique.fr/moodle/

Inscription to the course necessary:
https://docs.google.com/forms/d/11ZcoKzVenCEcVtgbxhuJ05CGBb2JrBWbPUoqQ7nOMHc/edit

Informative video for ALTEGRAD 2021:
https://www.dropbox.com/s/lcans4r4mlsryux/ALTEGRAD_VAZIRGIANNIS_2021-09-06.mov?dl=0

Prerequisites : Good level in Machine and Deep Learning, basic graph algorithms, python programming

Introduction to Computer Vision

Instructor: Alasdair Newson

Credits: 3 ects

Grading: lab work evaluation

Language: English

Syllabus: Image processing and Computer Vision are among the most common applications of data science. This is due to the fact that images and videos are particularly complex, and live in high dimensions. Therefore, tools to analyse and understand these objects are of great use to any data scientist studying them. These tools, which have existed before the advent of deep learning, look at images on a local scale and a global scale. The local scale corresponds to low-level information such as the colour of a pixel or the gradient in an image at a certain point. This is what we call image processing. Image processing is a pre-requisite for computer vision, which corresponds to understanding an image on a global level. In this case, we will be interested in detecting and understanding objects (faces for example) in images, to see what is happening in a scene. Such objects can be seen as collections of pixels, thus the more global viewpoint.

In this course, we will look at some of the classical tools to understand images. Firstly, we will go over basic sampling theory, used in signal processing. We will then look at filtering, radiometry, colour spaces and mathematical morphology. In the second lesson, we will look at computer vision, that is to say high level understanding of images. We will look at segmentation, basic feature detection (corners, lines etc), object detection and motion estimation. Finally, we will look at some more advanced topics in computer vision and image processing.

We will implement the algorithms seen in the lesson in lab works, carried out in the python programming language, with jupyter notebooks.

Format: 5 sessions

Course content:

- image sampling and filtering
- histogram equalisation
- image restoration
- image segmentation
- motion estimation

Programming language: Python, in the form of Jupyter Notebooks

Introduction to Reinforcement Learning

Instructor: Erwan LE PENNEC

Credits: 3 ects

Grading: Project based on a research article.

Language: English

Syllabus: This 20-hour course provides an introduction to reinforcement learning. It is based on the new edition of the book "Reinforcement Learning: An Introduction" by R. Sutton and A. Barto. Barto (available online at http://incompleteideas.net/book/the-book-2nd.html).

Outline:
1. Introduction to reinforcement learning and Markov decision processes
2. The bandit case
3. Tabular methods: prediction by dynamic programming, Monte Carlo method and TD Learning
4. Planning and learning for tabular methods
5. Approximate methods: prediction, planning and learning

Partially Observed Markov Chains in Signal and Image

Instructor: Wojciech PIECZYNSKI

Credits: 3 ects

Grading: Quiz, project based on a research article

Language: French

Syllabus:

Law and Ethics of Artificial Intelligence

Instructor: Winston Maxwell

Credits: 3 ects

Grading:

Language: English

Machine Learning Business Cases

Instructors: Paul Désigaud & various assistants (from Wavestone)

Credits: 4

Grading: The grade will be based mainly on a team project taking place along the course. A few quizzes & short activities will also be taken into account.

Numerus clausus: 40

Language: English

Syllabus :
Objectives: Discover how we do Machine Learning in companies, through concrete business cases, from their feasibility study to their industrialization.

- Consolidate theoretical knowledge and apply it to solve a business case.
- Learn how to design and develop ML applications.
- Learn how to share, valorize and restitute results to stakeholders.
- Learn how to work in a team.

Course description**:**
The course will consist of presentation sessions (data science theory and presentation of projects led for big companies) and labs (group work on provided use cases and data, preparation and presentation of the results).

Operations Research and Big Data

Instructor: Zacharie Arles

Credits: 3 ects

Grading: Examen écrit et projet

Language: French

Syllabus: La Recherche Opérationnelle (R.O.) est la discipline des méthodes scientifiques utilisables pour élaborer de meilleures décisions. Elle permet de rationaliser, de simuler et d'optimiser l'architecture et le fonctionnement des systèmes de production ou d'organisation. La R.O. apparaît comme une discipline-carrefour associant les mathématiques, l'économie et l'informatique.

Les apports de la R.O. sont visibles dans les domaines les plus divers : de l'organisation des lignes de production de véhicules à la planification des missions spatiales, de l'optimisation de portefeuilles bancaires à l'aide au séquençage de l'ADN ou à l'organisation de la couverture satellite des téléphones portables…

Tous ces problèmes sont de nature discrète ou combinatoire. Si l'existence d'une solution optimale est en général triviale, sa recherche de manière énumérative, même effectuée par les ordinateurs les plus puissants, pourrait demander plusieurs siècles de calcul.

L'objectif de ce cours en deux parties est de présenter les bases de la recherche opérationnelle et ses liens grandissants avec les sciences des données.

La partie 1 présente des problèmes classiques de théorie des graphes, l'algorithme du simplexe et le branch-and-bound. La partie 2 présente des liens entre la recherche opérationnelle et les sciences des données tel que l'intérêt apporté par la résolution exacte de problème pour obtenir des classifieurs à la fois performants et interprétables.

Prerequisites : Aucun pour la partie 1. Pour la partie 2, il est nécessaire d'avoir suivi un cours de recherche opérationnelle (que ça soit la partie 1 ou un autre) : théorie des graphes, simplexe, branch-and-bound.

Cloud Data Infrastructure

Instructor: Nicolas Travers

Credits: 3 ects

Grading: Project on real use cases, 5 deliveries

Numerus clausus: 40

Language: English

Syllabus: This course aims at describing how to model and distribute data efficiently in a distributed infrastructure dedicated to large-scale data management. The students will study techniques of denormalization, optimization of queries via modeling, sharding and indexing, information system constraints (consistency, persistence, distribution, etc.), and compare existing solutions and their specificities. The course will allow students to integrate the NoSQL eco-system for data management and to make relevant choices of adequate data infrastructure according to their own query needs.

Main themes :
- Learn data distribution/replication strategies & Compare replication algorithms (HDFS, DHT, GridFS),
- Distributed concurrency and consistency issues & algorithms (CAP, PACELC, 2PC, Paxos)
- Data modeling in a distributed context (NoSQL data families, cloud cost models)
- IS conception process over the Cloud

Recommended readings:
- T. Özsu, P. Valduriez, Principles of Distributed Database Systems, **4th Edition**, Springer (2020)
- S. Abiteboul, I. Manolescu, P. Rigaux, M.-C. Rousset, P. Senellart, « Web Data Management ». Cambridge University Press. http://webdam.inria.fr/Jorge/
- F. Abdelhedi, A.A. Brahim, F. Atigui, G. Zurfluh, "MDA-based Approach for NoSQL Databases Modelling". In: DAWAK 2017. pp. 88–102 (2017)
- A. Chebotko, A. Kashlev, S. Lu, "A Big Data Modeling Methodology for Apache Cassandra". In: IEEE International Congress on Big Data 2015. pp. 238–245. IEEE (2015)
- J. Mali, F. Atigui, A. Azough, N. Travers, "ModelDrivenGuide: An approach for implementing NoSQL schemas". In: DEXA 2020: pp. 300-310
- A. Fox, E.A. Brewer, "Harvest, Yield, and Scalable Tolerant Systems". In: HOTOS 1999. pp. 174–178. IEEE (1999).
- F. Atigui, A. Mokrani, N. Travers, "DataGuide : une approche pour l'implantation de schémas NoSQL". In: EGC 2020. pp. 407-408
- C. du Mouza, N. Travers, "Relevant Filtering in a Distributed Content-based Publish / Subscribe System". NoSQL Data Models: Trends and Challenges vol. 1, pp. 203–244 (2018)
- R. Behmo, N. Travers, "Maîtriser les bases de données NoSQL", https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql

Prerequisites: Database (E/R, SQL), basic knowledge in database optimization, UML class diagrams and notions in networks.

Auction Theory and In Practice

Instructor: Vianney Perchet (ENSAE)

Credits: 3 ects

Grading: Final exam.

Language: English

Syllabus: Billions of auctions are run everyday, in the amazingly huge online advertisement market. They require a complete knowledge of the different mechanism, how to improve them using past data and how to learn « good/reasonable/optimal » strategies.

During the different lectures, we will first introduce the general concept of mechanism design, and especially auctions (1st/2nd price, combinatorial, VCG) that are or can be used in practice. The main questions are their approximation, optimisation and learning based on past data in a dynamical setting. We will introduce and study the main classical tools with reminders on statistical theory, multi-armed bandits and online algorithms.

This course will be at the intersection of mathematics (statistics, optimization), computer science (complexity, approximation), and economics (strategies, equilibria and applications). Yet it should not have strong prerequisites. The evaluation will be a written exam.

At the end of this course, students will be able to :

1. Compute optimal strategies and design auctions
2. Find the sample complexity of approximate mechanisms

Prerequisites: Some knowledge of game theory would be preferable, but not compulsory. I will provide exercice sheets if needed.

References:

KRISHNA, Vijay. Auction theory. Academic press, 2009.
ROUGHGARDEN, Tim. Twenty lectures on algorithmic game theory. Cambridge University Press, 2016.
NEDELEC, Thomas, CALAUZENES, Clément, EL KAROUI, Nourreddine, PERCHET, Vianney. Learning in repeated auctions, https://arxiv.org/abs/2011.09365, 2020
NISAN, Noam, ROUGHGARDEN, Tim, TARDOS, Eva, *et al.* Algorithmic Game Theory

Deep Learning Advanced

Instructor: Edouard Oyallon

Credits: 3 ects

Grading: This class will be validated via a project presentation that will be based on a research article. The students will have to work by group of 2 up to 3.

Language: English

Syllabus: The objective of this class is to discuss several existing theoretical and practical results about deep neural networks. Those later consist in a cascade of linear and non-linear pointwise operators that are typically used for regression or generative tasks. They are now standard in many machine learning applications because they lead to outstanding performances. However, they are poorly understood from a theoretical perspective and they require many ad-hoc engineering tricks to be successfully trained. This class proposes to discuss several recent results, in simplified settings, through the lens of signal processing tools. Each lecture will last approximatively 1h30 and will be followed by a lab of 1h30, using potentially pytorch.

The outline of the lectures will be as follow:

1. A review of functional analysis and harmonic analysis tools which will be used

2. Group invariance and Neural Networks (Harmonic analysis for Machine Learning, Scattering Transform [4], ...)

3. Graph Convolutional Networks: from Euclidean to non-Euclidean data [2] (Wavelets on graphs, rediscovering standard Convolutional Neural Networks, ...)

4. Neural Networks from the Reproducing Kernel Hilbert Space perspective (Convolutional Kernel Networks, Neural Tangent Kernel, Lazy training [3], ...)

5. A case study: 1-hidden Layer Neural Networks (landscape, optimization of overparametrized Neural Networks, non-linear approximations, ...)

6. Deeper Neural Networks (complexity [1], Lipschitz-stability, ...)

References:

[1] Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally normalized margin bounds for neural networks. In Advances in neural information processing systems, pages 6240–6249, 2017.
[2] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine, 34(4):18–42, 2017.
[3] Lenaic Chizat, Edouard Oyallon, and Francis Bach. On lazy training in differentiable programming. In Advances in Neural Information Processing Systems, pages 2937–2947, 2019.
[4] Stéphane Mallat. Group invariant scattering. Communications on Pure and Applied Mathematics, 65(10):1331–1398, 2012.

Introduction to Time Series

Instructor: François Roueff

Credits: 3 ects

Grading:

Numerus clausus: 18

Language: English

Syllabus:

A Mathematical Introduction to Compressed Sensing

Instructor: Guillaume Lecué (ENSAE)

Credits: 3 ects

Grading: a well commentated python notebook report + defense.

Language: English

Syllabus: The objective of this course is to study several problems in high-dimensional statistics in order to identify three fundamental ideas of this theme that can be applied to many other data science problems. We briefly state these principles:

1) Many real world data belong to high-dimensional spaces in which classical statistical methods are inefficient (cf. the 'plague of the dimension'). Nevertheless these data are for the most part structured. So much so that the ``real'' dimension of the problem is no longer that of the ambient space but rather that of the structure which contains the useful information of the data. We speak of structured or parsimonious data in those cases. The construction of bases or dictionaries allowing to reveal the structures of these data is an important component of high dimensional statistics.

2) At first sight, the search for these low-dimensional structures seems to require the launching of a combinatorial search in a high dimensional space. However, such procedures cannot be used in practice. An important component of nowadays statistics is then to propose and analyze algorithms that can be implemented even in high dimensional spaces. For this purpose, an approach has received particular attention: the convex relaxation coupled with the toolbox of convex optimization;

3) Finally, the third component is the role played by randomness in high dimensional statistics. It turns out that structures are generally revealed by random objects and that, until now, we do not know how to exhibit these structures with deterministic measures as efficiently as the random ones.

A course in high-dimensional statistics can therefore cover several areas of mathematics including approximation theory, convex optimization and probability. In this course, we will mainly study the algorithmic and probabilistic aspects of this theory. Approximation theory will only be briefly discussed through the example of images processing.

This course will address the paradigm of high dimensional statistics mainly around three themes:

1) Compressed sensing: problem of exact and approximate reconstruction of a high dimensional vector from a small number of linear measurements of this vector knowing that it has a small support;
2) matrix completion: how to complete a matrix from the observation of a small number of its entries knowing that this matrix is of low rank; we will apply a matrix completion strategy to construct recommendation systems.
3) community detection in graphs: find high density subgraphs in 'large' graphs.

We therefore approach the problem of high dimensional statistics through three key objects/types of data: high-dimensional but parsimonious vectors, high-dimensional but low-rank matrices and finally, high-dimensional graphs whose nodes are organized in communities. The Compressed Sensing problem will be used as the main pedagogical vector for learning the three key ideas of high dimensional statistics mentioned above. We will therefore devote 4 sessions to it. Then we will devote 1 session to the matrix completion problem and 2 sessions to the communities detection problem.

From a technical point of view, we will mainly use ideas from probability theory and convex optimization. All practical sessions will be using Python and libraries such as cvxopt or cvxpy. Upon completion of this course, students should be able to

1) identify the computational properties of certain optimization problems, in particular, identify computationally difficult problems,
2) find convex relaxations for these (non-convex) optimization problems
3) understand the role of randomness for these problems, in particular to construct adequate measurement vectors
4) use compressed sensing, matrix completion and community detection problems as benchmark problems
5) build algorithms to approximate the solutions of convexified optimization problems.

Deep Learning II

Instructor: Yohan Petetin

Credits: 3 ects

Grading: Graded practical sesssion or project

Language: English

Syllabus: Ce cours présente les architectures de type "réseaux de neurones profonds" qui ont permis, ces dernières années, de grandes avancées pour des problématiques de classification, de prédiction ou de détection. Il s'agit de la suite du cours 'Deep Learning I' de la première période et ce cours se focalisera avant tout sur les fondements et la compréhension théorique de modèles dits **génératifs** tels que les "Restricted Boltzmann Machines (RBM)", "Deep Belief Network (DBN)", "Generative Adversarial Network (GAN)" ou encore les "Graph Neural Networks" qui permettent le traitement de graphes de données. Les étudiants seront également amenés à implémenter ces modèles dans le langage de leur choix.

Planning

- Rappel des méthodes d'échantillonnage de type Monte Carlo (principe, échantillonneur de Gibbs);
- Restricted Boltzmann Machine (Définition, propriétés, théorème d'universalité, apprentissage par l'algorithme Contrastive Divergence);
- Deep Belief Network (Définition, apprentissage);
- Auto-encodeurs variationnels (l'approche variationnelle, les auto-encodeurs);
- Generative Adversarial Network (Principe, universalité, extensions);
- Graph Neural Networks (principe, applications)

Audio and Music Information Retrieval

Instructor: Geoffrey Peeters, Gael Richard

Credits: 5

Grading: 30% labs/project + 70% written exam

Language: English

Syllabus: Audio and Music information retrieval is the interdisciplinary research field related to the extraction of semantic information from the audio signal; it allows the development of applications such as speech/music segmentation, recognition (environnemental sound classification, acoustic scene classification, musical instrument recognition), source separation, the estimation of specific music attributes (multi-pitch, tempo/beat, chord, structure), music identification by fingerprint (a la Shazam), cover detection pr au

totagging (into genre, mood) This course present the different facets of this field ranging from audio signal representation (Fourier, STFT, Constant-Q transform, audio features), music representation (pitch, chords, rhythm, structure), to pattern-matching and machine-learning models (DTW, HMM, generative/discriminative learning and deep learning).

Planning : 9 sessions of 4 hours + Exam

Mixed Effects Models : Methods, Algorithms, and Applications in Life Sciences

Instructor: Prague Mélanie

Credit: 3 ects

Grading: Final exam

Language: English

Syllabus : This course introduces methods for analyzing longitudinal repeated data characteristic in biomedical sciences. Mixed-effects models are the statistical tool of choice for modeling inter-individual variability in the evolution of longitudinal markers. In other words, the parameters of the model are considered to be random and to take a different value between the different individuals of a study. Furthermore, in life science, there is often a knowledge of the biological mechanism of a phenomenon. The latter can be described using behavioral models generally described by systems of ordinary differential equations (ODE). In this course we propose to develop mechanistic dynamic models where the trajectories are described by ODEs with mixed effects models on the parameters (a particular type of non-linear mixed effects model).
Pharmacokinetic-pharmacodynamic (PKPD) models are an example of mechanistic models that are extremely useful and used to study the effect of a drug among individuals in the same population. PK models describe how a drug is absorbed into the body, distributed and then eliminated, whereas PD models describe the course of the disease and the effect of the drug on the body. However, we will see that this type of approach has applications in other fields such as toxicology, agronomy, virology, immunology or cell biology.

Objectives:
The objective of this course is to understand how to define and use a mechanistic model. Through examples (PKPD, viral dynamics,...) we will study the methods and algorithms used for mixed-effects models: methods of parameter estimation, model construction, validation and selection. R and the Monolix software (http://lixoft.com/products/monolix/) will be (widely) used in the course.

References:

- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.
- Verbeke, G. and Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. Springer.
- Commenges, D., & Jacqmin-Gadda, H. (2015). *Dynamical biostatistical models* (Vol. 86). CRC Press.
- Lavielle, M. (2014). *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press.
-

NLP

Instructor: François Yvon (CNRS)

Grading: Final quizz.

Numerus Clausus: 50

Language: English

Course description: This class is an introductory course to the domain of Natural Language Processing (NLP). It is meant to provide a high-level overview of this vibrant field, which is evolving at a fast pace owing to the recent advances of deep learning models; it also covers some of the main algorithmic development notably aimed to process structured linguistic data such as syntactic parse trees or semantic graphs, as well as the deep neural architectures that are used to learn numerical representations for words and phrases; the course finally includes a glimpse at its most recent developments aimed at developing NLP systems in a multilingual contexts.

Main themes :

- Statistical NLP: A brief retrospective
- Words and the Lexicon
- The art of language modeling
- The essence of NLP: Models for Structured Data
- Shallow Semantics and Representation Learning
- Multilingualism and Machine Translation

Recommended readings:

- Jacob EISEINSEIN. Natural Language Processing. The MIT Press, 2019
- Yoav GOLDBERG. Neural Network Methods for Natural Language Processing. Morgan & Claypool Publishers. 2017. 287 pages. ISBN 978-1-62705-298-6

also:

- Julia Hirschberg and Christopher D. Manning (2016) Advances in natural language processing, Science Magazine.

Prerequisites: Basic statistics and optimization, formal language theory (automata and grammars) is a big plus.

Grading: Final quizz.

Applied Deep Learning with Python

Instructor: Charles Ollion et Olivier Grisel

Credits: 5 ects

Language: English

Syllabus:

Big Data and Insurance Project

Instructor: Denis Oblin

Credits: 3 ects

Grading: business case and oral presentation by team

Language: This course will be given in english or french with lecture material in french.

Syllabus: the objective of this course is to provide an insight of a specific industry and understand how data is leverage in this context. The course mix présentations (on insurrance industry, uses cases, project management insight) and testimony from insurance actors : established insurance company or start ups. You will have to identify a possible usecases and promote it to c-levers :

Course overview:
1) presentations:

- insurance industry insight
- how to find the question ? (not the easy part !)
- project management
- use cases & insurtech

2) 2-3 live testimony (to be updates), in french

Structured Data

Instructor: Florence d'Alché-Buc

Credits: 3 ects

Grading: Project defense

Numerus clausus : 30

Language: English

Syllabus: Many real-world applications involve objects with an explicit or implicit structure. Social networks, protein-protein interaction networks, molecules, DNA sequences and syntactic tags are instances of explicitly structured data while texts, images, videos, biomedical signals are examples with implicit structure. The focus of the course is solving learning and prediction tasks under this complex/structured assumption.While in learning and prediction problems the case of structured inputs has been investigated for about three decades, structural assumption on the output side is a significantly more challenging and less understood area of statistical learning. This course provides a transversal and comprehensive overview on the recent advances and tools for this exploding field of structured output learning, including conditional random fields, max margin approaches, surrogate approaches as well as end-to-end deep learning. These approaches illustrated on real applications will allow to cope with complex problems such as music transcription, question-answering, automatic captioning, molecule identification.

Tail Events Analysis: Robustness, Outliers and Models for Extremes

Instructor: Anne Sabourin, Pavlo MOZHAROVSKY

Credits: 3 ects

Grading: written exam

Language: English

Syllabus: Analysis of events in the tail of the distribution constitutes an important topic in statistics. The aim of the course is threefold: construction of the estimators less sensible to contamination of the data, identification of the outlying observations, and modeling of extreme events. The course starts with the introduction to robust statistics and measures of robustness where the concepts of the influence function and of the breakdown value are given. Then, the simplest univariate robust estimators of location, scale, and skewness are regarded which behave consistently even if the data is contaminated. These are further generalised to robust estimators of multivariate location and scatter (Stahel-Donoho estimator, minimum covariance determinant estimator, S-estimators, MM-estimators). Robust regression as well as PCA estimators are also considered. An important notion in robust statistics—data depth—is then introduced for the multivariate and functional framework. Presentation of the concept of the data depth function is followed by studying most important depth notions such as Tukey or projection depth, and (multivariate) functional depths. The regarded above material is then applied to detection of outliers in multivariate and functional data, as well as the cell-wise outliers. Finally an introduction to extreme values analysis is given. Here, extremes are defined as the largest values of a considered dataset. Extreme value theory suggests natural models for block-maxima and excesses above large thresholds, which give rise to estimates of quantities of major interest for risk management, such as high quantiles, large return levels, or tail probabilities outside the range of observed data. This aims on providing guidelines to the students for applying extreme value models to answer such practical questions.

Format: 6 × 3.5 hours + exam

Planning:

Week 1: Introduction to robust statistics. Measures of robustness: influence function, breakdown value. Univariate robust estimators of location, scale, skewness. Multivariate location and scatter estimators. Robust regression and robust PCA. Reading:Rousseeuw and Leroy (1987), chapters 1–3; Huber and Ronchetti (2009), chapters 1, 3, 8, 11; Wilcox (2016), chapters 1–3, 10; Rousseeuw and Driessen (1999); Hubert et al. (2005); Cornillon et al. (2012) for (basic) R programming.

Week 2: Lab session I. Univariate robust estimation. Robust multivariate estimation: projection pursuit, minimum covariance determinant estimator. Robust regression and ROBPCA.

Week 3: Data depth. Statistical data depth function: definition and properties, chosen notions. Data depth in infinite-dimensional setting. Identification of multivariate (row-wise and cell-wise) and functional outliers. Reading: Becker et al. (2013), chapters 1, 2, 4; Wilcox (2016), chapter 6; Tukey

(1975); Donoho and Gasko (1992); Zuo and Serfling (2000); Hubert et al. (2015); Rousseeuw and Bossche (2018).

Week: Extreme value statistics. The one dimensional case, distribution of maxima of large datasets, excesses above high thresholds. Inference methods (tail index estimation, block maxima, Peaks-over-thresholds). Case studies. Reading: Coles et al. (2001); Beirlant et al. (2006); Resnick (2013); De Haan and Ferreira (2007) .

Week : Multi-dimensional setting. Regular variation. Exponent and angular measure. Reading: Resnick (2007, 2013)

Week 6: Lab session II. Data depth: Tukey, projection, spatial depths, functional depths, applications to outlier detection. Extreme value analysis: Return levels, probabilities of failure.

References:

Becker, C., Fried, R., and Kuhnt, S. E. (2013). Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather. Springer, Berlin–Heidelberg.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). Statistics of extremes: theory and applications. John Wiley & Sons.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). An introduction to statistical modeling of extreme values, volume 208. Springer.

Cornillon, P., Guyader, A., Husson, F., Jegou, N., Josse, J., Kloareg, M., MatznerLober, E., and Rouviére, L. (2012). R for Statistics. Chapman and Hall/CRC, New York.

De Haan, L. and Ferreira, A. (2007). Extreme value theory: an introduction. Springer Science & Business Media.

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. The Annals of Statistics, 20(4):1803–1827.

Huber, P. J. and Ronchetti, E. M. (2009). Robust Statistics. Second Edition. John Wiley & Sons, Hoboken. Hubert, M., Rousseeuw, P. J., and Branden, K. V. (2005). Robpca: A new approach to robust principal component analysis. Technometrics, 47(1):64–79.

Topics in Stochastic Filtering, Information and Control

Instructor: Daniel Clark

Credits: 3 ects

Grading: The course will comprise of 15h lectures lectures and 15h tutorial and practical work. The assessment will be 30% coursework and 70% exam.

Language: English

Syllabus: Methods for estimating multiple objects from sensor data are in increasing de- mand and are critically important for national security. For example, the in- creasing use of space for defence and civil applications makes it imperativeto protect spacebased infrastructure. Advanced surveillance capabilities are needed to be able to identify and monitor activities in earth's orbit from avariety of different sensing platforms and modalities. There have been a number of important innovations in multitarget tracking and multisensor fusion in recent years that have had significant international impact across different application domains. In particular, the suite of mathematical tools, such as point process models, have been developed specifically to enable such innovations. Considering systems of multiple objects with point process models adopted from the applied probability literature enables advanced models to be constructed in a simple way. This course draws together mathematical concepts from diverse domains to provide a strong grounding for developing new algorithms for practical applications. This course will investigate mathematical concepts in multiobject estimation to enable prospective researchers to better understand and contribute to innovations in this field. The goal is to develop a broad mathematical perspective for mathematical modelling for multi-object estimation and explore the literature in spatial statistics and point processes to aid new advances in sensor fusion forthe development of future technologies for autonomous systems.

Course content: The topics have been selected to cover the fundamental topics required for the development and implementation of practical algorithms for multi-sensor fusion. The course will cover fundamental mathematical topics, in estimation theory, information theory, and point process theory as follows.

- Bayesian filters: Kalman filter, extended Kalman filter, unscented Kalman filter, sequential Monte Carlo (particle) filtering, Gaussian mixture filtering
- Performance bounds and analysis: Fisher information, Cramer-Rao lower bound, consistency and bias.
- Topics in combinatorics: generating functions, Bell polynomials, partitions.
- Topics in functional calculus: differentials, functional derivatives, generating functionals
- Point process statistics: the intensity function, covariance and correlation, moments and cumulants
- Point process descriptions: the probability generating functional, the Laplace functional

- Topics in multi-target tracking: modelling and derivation of point process filters, application with Gaussian mixture and particle filters,
- Practical applications: simultaneous localisation and mapping (SLAM), tracking multiple targets and camera calibration, distributed multi-sensor multi-target tracking.
- Topics in information: Shannon entropy, Kullback-Leibler divergence, Renyi entropy, mutual information, channel capacity.

Optimal Transport: Theory, Computations, Statistics, and ML applications

Instructor: Marco Cuturi

Credits: 3 ects

Grading:

Language: English

Syllabus: 8 séances de cours + 4 séances de TD = 18 heures

3 hours: theory

- Monge and Kantorovich Problems, duality in OT.
- 2-Wasserstein geometry and the Brenier theorem. Applications to transport between Gaussians
- Transport in 1D
- Wasserstein space: JKO flow.

7.5 hours: computations and statistics

- Algorithmic overview: network flow solvers in the discrete world, Benamou-Brenier formula in the PDE world.
- Statistical results and the curse of dimensionality
- Regularized approaches to compute optimal transport.
- Differentiation of Optimal transport: unrolling / implicit.
- Unbalanced generalizations, quadratic extensions (GW).

2 TDs:

- 1D transport, transport between Gaussians
- Network flow solver type algorithms + SInkhorn algorithm

3 hours: applications

- Handling measures with the Wasserstein geometry: barycenters, clusters
- Wasserstein GANs
- Applications to Biology (cell pathways) and NLP (alignment of multilingual corpora)

2 TDs:

- Differentiation of Sinkhorn's algorithm
- Transport between high-dimensional point clouds.